



## Electronic Outlook - A Comparison of Human Outlook with a Computer-vision Solution

**Blanke, Mogens; Hansen, Søren; Stets, Jonathan Dyssel; Koester, Thomas ; Brøsted, Jesper E.; Maurin, Adrian Llopert; Nykvist, Nicolai**

*Link to article, DOI:*

[10.13140/RG.2.2.26172.08325](https://doi.org/10.13140/RG.2.2.26172.08325)

*Publication date:*

2019

*Document Version*

Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*

Blanke, M., Hansen, S., Stets, J. D., Koester, T., Brøsted, J. E., Maurin, A. L., & Nykvist, N. (2019). *Electronic Outlook - A Comparison of Human Outlook with a Computer-vision Solution*. Technical University of Denmark. <https://doi.org/10.13140/RG.2.2.26172.08325>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



# **Electronic Outlook - A Comparison of Human Outlook with a Computer-vision Solution**

**Mogens Blanke<sup>1</sup>, Søren Hansen<sup>1</sup>, Jonathan Dysse Stets<sup>2</sup>  
Thomas Koester<sup>3</sup>, Jesper E. Brøsted<sup>3</sup>,  
Adrian Llopert Maurin<sup>1</sup> & Nicolai Nykvist<sup>1</sup>**

<sup>1</sup> Department of Electrical Engineering, Automation and Control Group,  
Technical University of Denmark,  
Elektrovej Building 326, 2800 Kongens Lyngby

<sup>2</sup> Department of Applied Mathematics and Computer Science,  
Image Analysis and Computer Graphics, Technical University of Denmark,  
Richard Petersens Plads, Building 324, 2800 Kongens Lyngby

<sup>3</sup> FORCE Technology, Hjortekærsvej 99, 2800 Kongens Lyngby

Updated version of March 19, 2019

Research report prepared for the Danish Maritime Authority



Printed at DTU  
Kongens Lyngby, Danmark  
March 2019

© 2019 DTU and FORCE Technology  
This report can be freely distributed with proper reference to the source.  
No changes are permitted in the document.

**ISBN** 978-87-91184-07-9  
**DOI** 10.13140/RG.2.2.26172.08325

# Contents

<b>Contents</b>	<b>3</b>
<b>1 Summary</b>	<b>5</b>
<b>2 Introduction</b>	<b>9</b>
<b>3 Outlook for navigation</b>	<b>11</b>
3.1 Human outlook . . . . .	11
3.2 State of art . . . . .	13
3.3 Scope of this investigation . . . . .	15
<b>4 Electronic outlook</b>	<b>17</b>
4.1 Camera comparisons . . . . .	18
<b>5 Object detection and classification</b>	<b>23</b>
5.1 Image-based Object Detection . . . . .	23
5.2 Related work . . . . .	24
5.3 Mask R-CNN detection and classification . . . . .	25
5.4 Dataset . . . . .	25
5.5 Training . . . . .	27
5.6 Performance . . . . .	29
<b>6 Results</b>	<b>33</b>
6.1 Temporal Comparison . . . . .	33
<b>7 Discussion</b>	<b>37</b>
<b>8 Conclusions</b>	<b>39</b>
<b>Bibliography</b>	<b>41</b>



# Chapter 1

## Summary

Considering whether a temporarily unattended bridge could be allowed, Maritime Authorities wish to investigate whether sensor technology is available that, when seconded by sophisticated computer algorithms, is able to provide outlook with the same reliability and safety as that of the average human outlook.

This document report findings from a comparative study of human versus electronic outlook. Assessment of navigator's outlook is based on measurements with a wearable eye-tracker and areas of their visual attention are recorded on video. Simultaneously, a set of electro-optical sensors provides image-data as input to computer algorithms that detect and classify objects within visual range.

Ambient light conditions on the bridge prevented eye-tracking to disclose which objects on Radar and ECDIS screens caught attention of the navigator. The scope of this investigation was therefore limited to focus on the navigator's visual attention on objects. The report compares these eye-tracking measurements with object recognition made by camera recordings in the visual spectrum and subsequent computerized object classification.

The report deducts, from the observations of eye fixations, when the navigator became aware of a particular object. It analyses how the human observations compare with those of the technology solution. On the technology side, the report presents approaches to detection and classification, which appeared to be efficient in coastal areas with confined passages. The quality of outlook in different ambient light conditions is illustrated. The main findings are:

- Eye-tracking glasses were found useful to show fixations on objects at sea in daylight conditions.
- The computer-vision algorithms detects objects in parallel, the human does so sequentially, and the computer classifies objects in average 24 sec faster than the navigator has a fixation on the object. The deep learning algorithm trained in this study should, however, be improved to achieve better performance in some situations.
- The time between object detection and passage of own ship is adequate for making navigation decisions with both human and electronic outlook.

- Low-light conditions (dusk and night) are effectively dealt with by Long Wave InfraRed (LWIR) camera technology. LWIR shows objects as equally visible at day and night.
- Colour information from cameras is necessary to assist decision support and electronic navigation.
- A system for electronic outlook should employ sensor and data fusion with radar, AIS and ECDIS.
- Decision support based on electronic outlook should include object tracking and situation awareness techniques.
- Quality assurance and approval of machine learning algorithms for object classification at sea has unsolved issues. A standard vocabulary ought be available for objects at sea, and publicly available databases with annotated images from traffic in both open seas, near coast areas and rivers should be available in order for authorities to assess quality or approve navigation support based on machine learning methods.

## Abbreviations and acronyms

---

AIS	Automatic Identification System for marine vehicles
CNN	Convolutional Neural Net
ECDIS	Electronic Chart Display and Information System
GNSS	Global Navigation Satellite System
HFOV	Horizontal Field of View
JAI	Manufacturer of industry-grade cameras
LWIR	Long Wave InfraRed (8 - 14 $\mu m$ wavelength)
Mask R-CNN	R-CNN network with segmentation dedicated
NIR	Near InfraRed (800-1000 $nm$ wavelength)
Radar	RAdio Detection And Ranging system
R-CNN	Region based CNN
RGB	Red Green Blue color coding of digital image
RoI	Region of Interest used in image analysis
Teledyne Dalsa	Manufacturer of LWIR camera and others
Tobii® glasses	Eye-tracking glass from Tobii AB, Sweden

---



# Chapter 2

## Introduction

Look-out for navigation is the task of observing various objects which can have an impact on a ship's planned route and maneuvering capabilities, for example other vessels, buoys and land. If the outlook is a separate person on the bridge, observations are reported to the officer in charge who decide any remedial actions. The look-out is made using sight and aided by available technology such as radar, AIS and ECDIS systems. Development within camera technology and computer vision algorithms has provided an additional possible source for look-out. This report investigates the quality of this “electronic outlook” and compares with human look-out.

A survey of maritime object detection and tracking methods was recently (2017) published in the survey by [24], who emphasized that Radar, which is an IMO required instrument on merchant vessels, is sensitive to the meteorological condition and the shape, size, and material of the targets. They emphasize that radar data need to be supplemented by other situational awareness sensors to obtain safe navigation and collision avoidance. Commercial electro-optic sensors are available for several spectral ranges: visible (450-800 nm), near infrared, (NIR 800-950 nm) and long wave infrared (LWIR 8-14  $\mu m$ ). [28] investigated detectability of objects at sea seen from an aircraft in visible and LWIR spectral ranges.

A brief overview of the approach taken in this investigation was presented in [4].

This report first summarises the task of watchkeeping/lookout for navigation. Chapter 3 describes how human outlook is observed through measurements where a navigator wears eye-tracking glasses. Chapter 4 outlines the use of electro-optical and other sensors to provide electronic means to replicate the human observation of surroundings. Chapter 5 introduces to available technology for object detection and classification at sea where image processing and machine learning techniques are instrumental. Chapter 6 presents findings of this study from ferries in near-costal and shallow water navigation. Chapter 7 discusses the results, their limitations and perspectives, and Chapter 8 finally offers conclusions and outlines directions for further research.





# Chapter 3

## Outlook for navigation

The analysis of manual lookout/watchkeeping is based on a combination of observations on board several vessels in Danish waters. Camera recordings and mapping of objects at sea were conducted on measurement campaigns during 2018. Eye tracking measurements were conducted during the summer of 2018.

Background knowledge for this study include generic observations, made by the FORCE Technology in Lyngby, on board a large number of vessels during the period 2000-2018. The generic experience also includes observations from ship simulator exercises, as well as literature-based studies, [31], [33], and on general knowledge of maritime human factors. Eye-tracking glasses have been used in simulator context but the use on board a bridge seems to be a new undertaking.

### 3.1 Human outlook

The manual outlook serves several purposes: prevention of collision with other vessels, safe and efficient navigation in the sea, avoiding grounding and collision with either fixed or floating objects, optimal steering and manoeuvring in relation to waves, wind and visibility, general observation of the sea, including observation of not normal conditions like emergencies or pollution.

The look-out also involves an element of continuous self-monitoring of the efficiency of the look-out and adjustment when necessary, e.g.: in cases including reduced visibility, rain, disturbing sun light, observe far away objects in binoculars, stress condition or fatigue of the look-out.

Outlook is made by both sight and sound, which literally means that not only the eyes are used, but also the hearing is active. Attention is made to maneuvering, fog or emergency signals from other vessels and to relevant radio communication.

The outlook is just one among several tasks of the navigator on the bridge. Other tasks that require attention include, observation of the condition of engines and systems, handling of cargo and passengers, safety-related routines, communication internally on board the vessel and with external parties, management of staff and other administrative tasks, QA and documentation and handling of safety-critical situations on board. These supplemental duties imply that the navigator needs to share his/her attention among several tasks.



Figure 3.1: Screen-shot from eye tracking of the navigator onboard the Elsinore-Helsingborg ferry M/S Pernille

## Endogenous and exogenous driven visual attention

The look-out task involves both endogenous- and exogenous-driven activities.

Endogenous activities are visual attention controlled by the navigator himself on his own initiative and based on relevant knowledge and experience, such as observing navigational markings, sighting of land and watching out for other vessels.

Exogenous activities are those caused by an external event catching the attention of the navigator. For instance, the sight of a vessel that the navigator has not been looking for or signals by light, sound or radio. Everyday scenarios will typically be a combination of endogenous and exogenous look-out activities.

As an example of the activity of endogenous controlled visual attention by the navigator, look at the screen-shot in Figure 3.1.

In this case the ferry Pernille is crossing a traffic separation scheme in which inbound traffic could be approaching from port side. The visibility (for both navigator and radar) can easily be obscured by the Kronborg Castle headland, but the navigator knows from experience that larger ships with considerable speed might be approaching. Based on this experience the navigator keeps an eye on AIS information and on the headland area to stay prepared for possible traffic at port side. The visibility of the ferry Pernille from other vessels is likewise obscured by the same headland, thus making it a possible hazard to proceed without extra caution. The navigator's visual attention is not drawn to the area by a concrete vessel (which would have been an activity of exogenous visual attention) but directed there based on his experience and anticipation, which is an activity of endogenous visual attention.

When it comes to performing an outlook, it makes sense to distinguish between two types, namely the pure observations that do not requiring action and observations that requiring action, e.g. to follow COLREGS and ultimately to prevent a collision or grounding. The navigator's actions are often seen as a combination of several elements including signaling, steering and engine manoeuvres.

The decision to act or not to act depends on the navigator's interpretation of the information stemming from these look-out related behaviours:

1. Repeated observations to determine if there is a risk of collision and if so taking countermeasures
2. Repeated observations to determine if executed countermeasures (during 1) above) have the desired effect
3. General visual observation (watching) of nothing in particular, but often focused on the direction of the vessel and abeam/passed in relation to the progression of the navigation
4. Exogenous visual attention triggered by something turns up, possibly in combination with information on bridge instruments
5. Endogenous visual attention, i.e. the navigator expects upcoming, observable objects using e.g. sea charts (buoys), radar or AIS.

The above unfolding of human outlook reveals the activity as complex and multifaceted. It is therefore generally accepted in the maritime domain, that there is no unified, single, correct or "perfect" way to perform outlook. Rather there is an acceptable margin within which different navigator's different behaviours can unfold, while still fulfilling the purpose of a good-enough outlook.

## 3.2 State of art

In the maritime context, the use of eye tracking as means to examine the visual attention of ship navigators is nothing new. At least not when it comes to the use of eye tracking in simulation environments. [3] investigated the operators' foci of attention during simulated dynamic position operation. [2] examined the difference in attention-allocation comparing novice and expert navigators during use of the Conning Officer Virtual Environment, a simulation system developed to train shiphhandling. [2] concluded a clear link between the experts' superior shiphhandling performance and a "tight Attention-allocation pattern that focused only on the relevant areas of interest. Novices' Attention-allocation patterns were highly scattered and irregular" (p. xviii). [22] and [27] focused on evaluating and improving the training of navigators using eye tracking data and [23] suggested using (stationary) eye tracking to determine or monitor the level of fatigue in the boat driver with the purpose of enhancing situation awareness. [15] used eye tracking data examination to suggest improvement of usability design on the ships' bridge layout and in the software's graphical user interface on a maritime navigation display. [14] also investigated eye tracking data in the pursuit of a recommendable optimal *visual scan pattern* for navigators aiming to mitigate the mental workload needed to monitor the increasing amount of technology used at the maritime ship bridge.

[11] performed a somewhat rare example of an investigation using eye tracking during actual, real life navigation. They investigated gaze behavior data from 16 experienced and novice boat drivers during high speed navigation and concluded that novices looked more at objects closer to the boat while experts looked more at

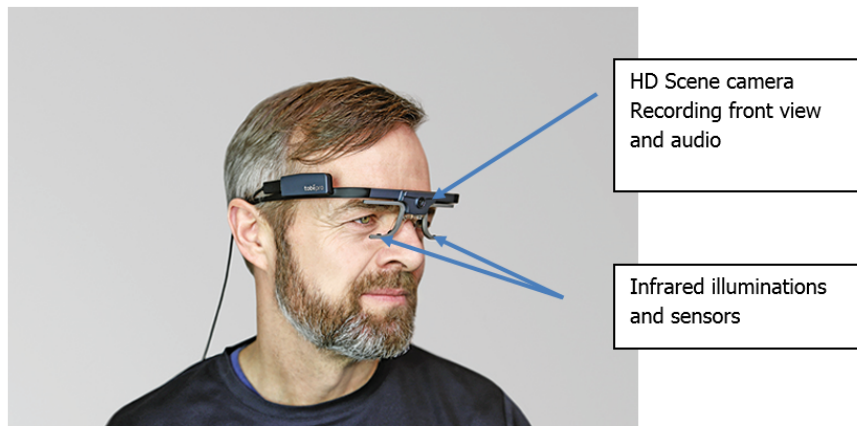


Figure 3.2: Tobii eye tracking glasses. *Photo: FORCE Technology 2018*

things far from the boat. Also, novice boat drivers were more focused on electronic displays, while the experts were focused mostly outside the boat and “further used the paper-based sea chart to a larger extent than the novice drivers” (p 277).

The methodology of using eye tracking devices in real life maritime situations is not often seen, thus making this study quite unique.

### Eye tracking technology applied in this investigation

The eye tracking data was collected using Tobii Pro Glasses 2 ([1]), which is a lightweight wearable technology. It consists of two units: a head unit (the glasses) connected via a HDMI cable to a belt clip recorder unit.

The head unit has a scene camera recording the wearer’s front view (including audio) and the frame has infrared illuminators and sensors installed thereby using the eye tracking technique Corneal reflection (dark pupil). The belt clip unit holds a SD card for recording data, operates on rechargeable batteries and is Wi-Fi controlled through PC-based software (in this case iMotions). This setup makes it very easy for the person wearing the eye trackers to freely move around on the ship and due to the non-invasive design, most subjects easily forget they are even wearing them while performing their job. Additional specifications are shown in the table below, adapted from the Tobii Pro Glasses 2 User’s Manual (2018, p. 40).

Based on the recording from the scene camera and the associated eye tracking data, the iMotions software (current version 7.1) produces a video showing what was in the wearer’s field of view during the recording (a 1st person perspective replay), including a graphical overlay: A yellow dot indicating where in the field of view the person was looking at any given time. The software can be set to illustrate fixations by either increasing size of the yellow dot or color change hereof. A fixation is defined as a period (100 ms or more) in which the person’s eyes are locked toward a specific object (or location) in the field of view. Fixations are excellent measures of visual attention. Fixations are directly related to cognitive processing [17] and during a fixation, the person is analyzing and interpreting information from the object or area of focus [22].



Figure 3.3: Eye tracking example.

The image in Figure 3.3, shows a single frame from replay of an eye tracking recording (MF Højestene). The yellow dot is the location of the navigator's fixation and the yellow line illustrates eye movements faster than 100 ms, referred to as saccades.

### 3.3 Scope of this investigation

In the course of using the Tobii® eye-tracking glasses on a bridge and iMotion® software, the following limitations appeared

- Objects at the Radar of ECDIS screens are tiny compared to outside objects and the fixation circle is too big to distinguish which object on a screen is catching attention
- Very high differences in light intensity from outside to instrument screens
- Slight drift between direction of view indicated by the eye-tracking and actual direction
- The Tobii glass camera is excellent for daylight conditions but is not sensitive enough for dusk or night conditions.

These technical limitations with eye tracking measurements restricted the scope of this investigation, and it was necessary to limit the scope to simply compare eye-tracking fixations on objects at sea with classification of the same objects made by the computer-vision based part of electronic outlook. Observation with eye-tracking was limited to daylight conditions.

The electronic outlook then focus on object detection and classification in daylight images. Machine learning techniques are explained and the quality of object recognition at sea is assessed from training and validation images. It is discussed why the usual softmax, precision and recall measures from deep-learning are not

sufficient as quality indicators for navigation purposes. Instead, we show how traditional statistical measures (true and false, positives and negatives) are employed at an object level and provide useful quality indicators for use in navigation.

This study has not employed means to disclose how the navigator interprets what he sees. The eye tracking glasses can determine where the navigator has had visual focus. The detailed recognition of objects and their behaviour are therefore not in the scope of this investigation. The study does not reflect on which actions the navigator could or should make.

# Chapter 4

## Electronic outlook

This chapter describes the instrument platform used for electronic outlook and it presents images available from the camera suite in different spectral ranges selected for the study.

The electronic outlook system in this comparison consists of 5 cameras, an FMCW radar, an AIS receiver, a GPS unit and a 6 degrees of freedom inertial measurement unit.

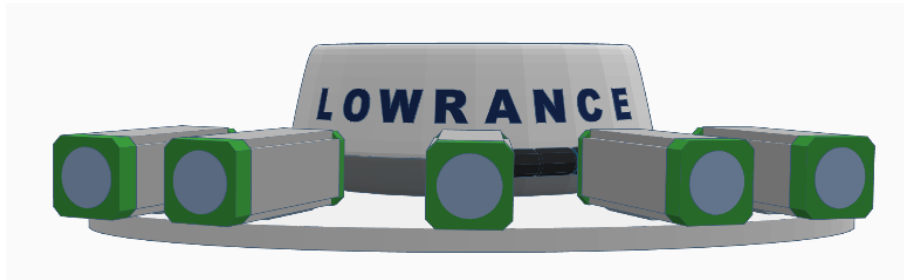


Figure 4.1: Sketch of the sensor platform. The five camera houses are looking forward in the ships heading. Camera units, CW-FM radar and GPS receiver plus IMU are fixed on the platform.

The vision system comprises,

- 2 color cameras (JAI GO-5000C), 5M pixel, resolution is 2560 x 2048, 12 bit, lens 55 HFOV. A polarizing filter is added.
- 2 monochrome cameras (JAI GO-5000M), 5M pixel, resolution is 2560 x 2048, 12 bit, lens 55 HFOV. NIR low pass filter (800-1000 nm can be added).
- 1 infrared LWIR camera, uncooled micro bolometer sensor (Teledyne Dalsa Calibir 640), resolution 640 x 480, 14 bit, lens has 90 deg HFOV.

The equipment is mounted on a forward facing stand on board the ferries. Object detection and classification algorithms are run as post-processing of the images. The platform is sketched in Fig. 4.1. The color and monochrome cameras have an overlap of 4 degrees. Fig. 4.2 shows the instrument platform mounted on a ferry in a narrow passage in the southern Funen archipelago.





Figure 4.2: Southern Funen archipelago. Sensor platform mounted beyond wheel-house / ship's bridge.

## 4.1 Camera comparisons

Measurements were made to assess camera properties in daylight, at dusk and at late evening to assess image properties in conditions of normal, low and very low ambient illumination.

### Daylight

Daylight conditions depicted by RGB camera recordings and monochrome NIR range photos are shown in Figures 4.3 and 4.4. High contrast is a visual challenge as visual images have 8 bit depth only, but the full 12 bit resolution of the cameras is maintained in the subsequent image analysis.



Figure 4.3: RGB camera with polarizing filters. South Funen archipelago. Ambient condition is daylight and good visibility



Figure 4.4: Monochrome camera with NIR filter  $800 - 1000nm$ . South Funen archipelago. Ambient condition is daylight and good visibility



Figure 4.5: LWIR camera  $8 - 14\mu m$ . South Funen archipelago. Ambient condition is daylight and good visibility

The LWIR image in Figure 4.5 is recorded at the same instant as the RGB and NIR images. The 14 bit resolution of this camera allows a temperature resolution of objects of  $0.006$  deg. When visualizing the LWIR image, we use range compression to zoom in on the temperature range that is relevant for the image. The LWIR is not sensitive to sun reflections in the water or in clouds. As seen, several vessels are recognizable in the LWIR image, which are less apparent in the RGB and NIR pictures, even though the LWIR has 4 times lower horizontal resolution than the JAI cameras.

### Dusk - 26 min after sunset

Figures 4.6, 4.7 and 4.8, show the RGB, NIR and LWIR images observed 26 min after Almanac sunset. Navigation lights on the approaching ferry are clearly visible on in the NIR image, and the heat silhouette very clearly visible in the LWIR image.

Disturbed by the glare from sunset, the RGB image will need contrast enhancement to distinguish the navigation lights on the approaching ferry.

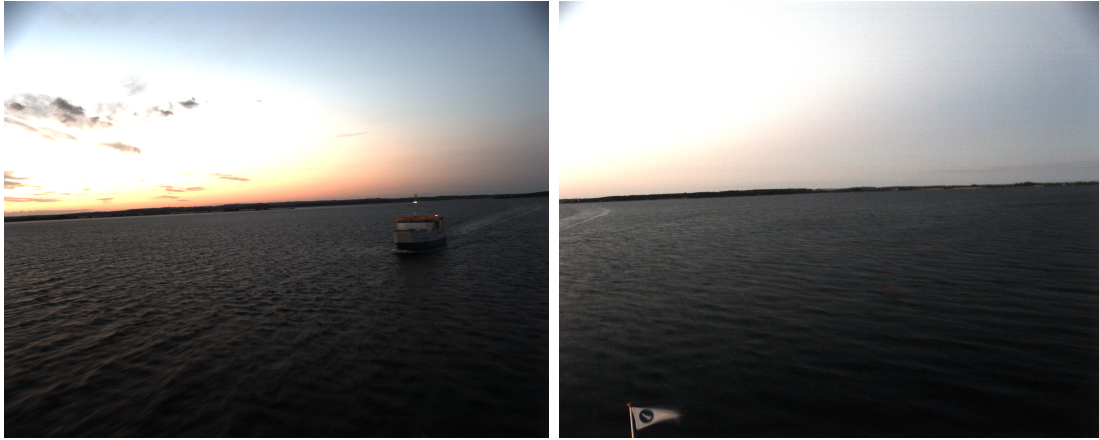


Figure 4.6: RGB camera with polarizing filters. South Funen archipelago at dusk



Figure 4.7: Monochrome with NIR filter. South Funen archipelago at dusk

### Dark - 60 min after sunset

Measurements were made with Cameras mounted with parallel optical axes, pointing forward in own vessel. Figure 4.10 shows crossing traffic, a ferry with illumination along the vessel and from cabins, saloons and deck, and a freighter behind with two mast lights and sparse deck illumination. The cargo vessel's lights could be mistaken for land/street illumination. Images from Long Wave Infrared Cameras, shown in Figure 4.9, are not sensitive to visual light, but only to surface temperatures of observed objects, and both of the crossing vessels are clearly distinguishable. The LWIR cameras have different horizontal field of view (HFOV),  $45deg$  for the FLIR camera and  $90deg$  for the Teledyne Dalsa. While the LWIR cameras clearly show signatures of other vessels, the red-white-green colour codes of navigation lights provide essential information that is visible in the RGB images, but not in the LWIR images.

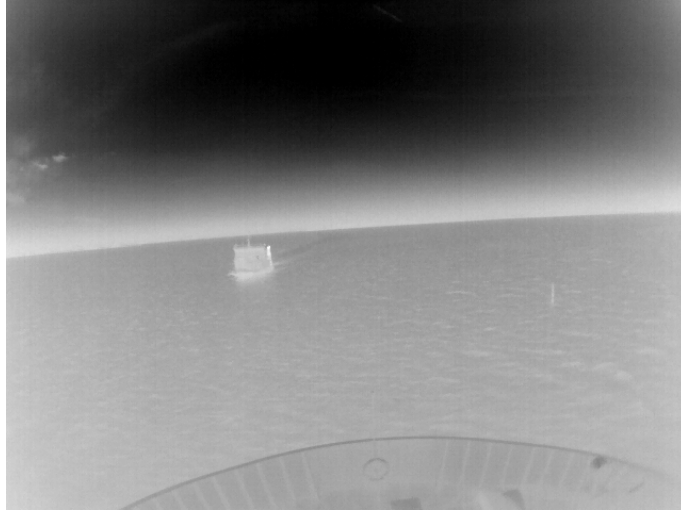


Figure 4.8: LWIR camera, south Funen archipelago at dusk



(a) Teledyne-DALSA LWIR ( $90^\circ$  HFOV)



(b) FLIR LWIR ( $30^\circ$  HFOV)

Figure 4.9: LWIR camera recordings,  $640 \times 480$  pixels. Oresund toward Elsinore. Ambient condition: dark but good visibility one hour after sunset





Figure 4.10: RGB image. Deck and cabin lights are clearly visible on the crossing ferry, a cargo vessel behind it is somewhat difficult to distinguish from lights at the coast line. Strait of Oresund toward Elsinore. One hour after sunset

# Chapter 5

## Object detection and classification

This chapter gives a brief overview of object detection and classification techniques employed in this study and provides a description of the choices made for sensor system design. We start by introducing the object detection methods and related works. Then we introduce learning-based object detection algorithm we use and the training data captured with the on-board camera equipment. Finally, we discuss the results and performance of our implementation.

### 5.1 Image-based Object Detection

Object detection and classification is the task of determining what is present in a given image. There can be multiple ways to implement a detector and classifier, and what method to use depends on the specific use case. The relevant types of object detection and classification outputs are described below to give an understanding of what results can be obtained from images:

**Image Classification** To determine what one or more classes that an image can belong to. Often, the result of image classification is a probability of the image belonging to certain classes from a predefined list. In a maritime environment, classes such as 'ferry', 'sailboat', 'buoy', etc. would be obvious examples. It is however not directly possible to tell what specifically in the image that caused the classification result.

**Object localization** To localize an object in a specific image with, e.g., a bounding box. With object localization, it is possible to tell where in the image a specific object is located as well as an approximation of its relative size in the image. In other words, we can get a coordinate of the position of the object in the image.

**Semantic segmentation** Is a pixel-wise segmentation of one or more objects in the image. This will provide more exact information, as only the pixels that belong to a certain object is classified. Consequently, it is possible to not only get the position of the object in the image, but also a good approximation of the object shape.

**Instance segmentation** Is similar to the semantic segmentation, but objects of the same class will be classified as different instances. This is advantageous when two of the same class are partially overlapping, they will be segmented separately. This enables the possibility to count the number of objects belonging to a class or prepares the base for tracking them individually in a series of images.

Many proposed solutions exist for all of the above-mentioned outputs. Recently, data-driven solutions, such as deep neural networks, have proved to give robust and accurate results but these require large sets of annotated training data. Annotations often have to be done manually, and especially pixel-wise annotations for semantic and instance segmentation requires accurate and therefore cumbersome annotation. Techniques that require less or no prior data also exist but tend to be less general than a learning-based approach. Since our system is operating near the coast, many types and sizes of boats and ships can appear in the images. Additionally, we can have both land and water as background. The following section provides an outline of some challenges for maritime environments along with related prior work.

## 5.2 Related work

Object detection, classification and tracking in a maritime environment is a well-explored area, and several previous works address this. Challenges include waves that can cause a rapid change in the frame of reference [10], sudden change of illumination and unwanted reflections from the water [5], and the possibility of poor weather conditions that reduce the range of sight. The survey papers [24], [21] mention a range of methods that deal with detection and classification in images of maritime environments. Horizon line detection and background subtraction appears to be effective for object detection [34], [32], [25]. Object tracking was shown by [7] to be efficient when using a scale-invariant feature transform (SIFT). Utilizing infrared and visible light images was analyzed in [24], and also thermal imaging seems to have the ability to provide information about objects on the water [19]. With recent progress in deep learning based segmentation and classification methods, visible light images is an obvious choice for object detection. Training data is available in the ImageNet [8] picture base, but images already exist and can provide a base for training. For specifically maritime environments [18] and [6] show that deep learning based methods are effective and annotated data from the maritime environment exist from harbour of Singapore [24]. This study has used training data collected from observations on board ferries in Danish coastal waters.

In our study, [30] investigated different networks for machine learning, including R-CNN and Fast R-CNN, which do not employ segmentation. The classification was of modest accuracy in general, but objects with many instances in the training data set, in this case buoys in Danish coastal areas, performed much better than average precision and recurrence. False positives observed by machine learning, e.g. confusing a cloud for a sail, could be remedied by combining deep learning with methods from classical computer vision. These were investigated in [25].

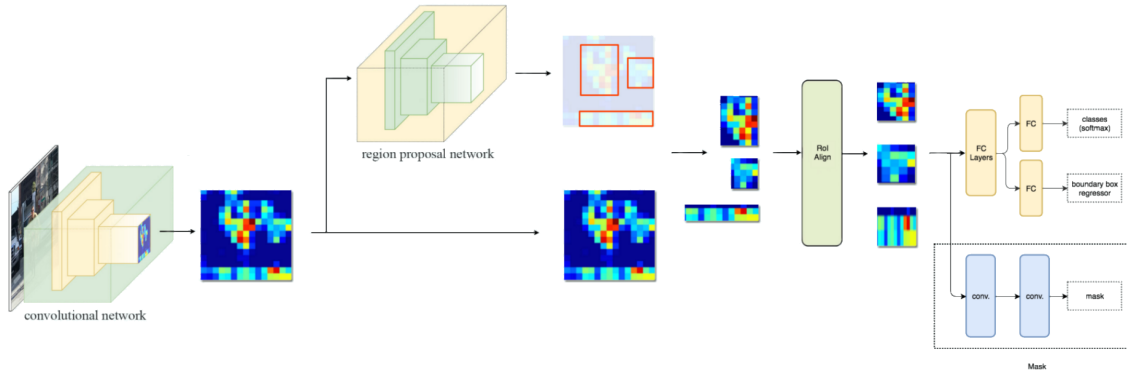


Figure 5.1: Mask R-CNN network. (From: [https://medium.com/@jonathan\\_hui/image-segmentation-with-mask-r-cnn-ebe6d793272](https://medium.com/@jonathan_hui/image-segmentation-with-mask-r-cnn-ebe6d793272))

### 5.3 Mask R-CNN detection and classification

Based on results of eye-tracking measurements, we decided to use the Mask R-CNN [16] on visible light images, which is able to perform pixel-wise segmentation of several classes. This section will give an introduction to the specific architecture of the Mask R-CNN implementation utilized for the visible light images. Objects that are within visual range of the cameras are detected and classified using a Convolutional Neural Network (CNN), also referred to as deep learning technology. The network architecture employed in this study to detect different objects in the maritime environment is Mask R-CNN [16], which has the novelty of not only being able to recognize and detect (bounding box) of several classes, but is also able to segment all instances of each one and create the corresponding binary mask at a pixel level. Mask R-CNN is the culmination of an architectural model that started with a Region-Based Convolutional Neural Network (RCNN) [13], followed by Fast-RCNN [12] and then Faster-RCNN [26]. The Mask R-CNN architecture, shown in Figure 5.1, was employed in this study. It has branch at the end of the CNN that predicts mask-segmentation at a pixel-to-pixel level. Misalignment and loss of data are avoided, and pixel level precision is obtained by a RoI-Align layer that uses bi-linear interpolation to remove quantization at the region of interest boundaries.

### 5.4 Dataset

In this study object detection is currently purely based on image data, so the following will give a basic description of the digital images produced by our vision system. We found that existing maritime image data-sets are not sufficient to cover the scenarios we encounter in our recordings. Consequently, we choose to manually label a small set of images to refine the Mask R-CNN. The annotation process will also be mentioned in this section.



## Image Format

We define a digital image as a 2-dimensional array of discrete values representing the pixels in the image. The number of horizontal and vertical pixels in the image is referred to as the resolution of the image and is a determining factor of how detailed the information of a given scene can be stored in an image. The amount of information stored per pixel is referred to as the bit-depth. In a grayscale image, a pixel commonly contains 8 bits of information which means that each pixel can take 256 levels of intensity. A higher bit depth allows storing more information in each pixel, as for example, a higher dynamic range. An image can consist of one or more channels, and an example of this is a colour image which contains 3 channels: red, green and blue respectively. This means that a colour image takes more storage space and consequently is slower to process than a grayscale image of the same resolution. On the other hand, it contains more information which can be valuable for some applications. In our case, the resolution the colour images are a quarter of the monochrome images, because they need to be GRBG demosaiced to produce a colour image from their raw format.

To be able to accurately detect objects in the images we acquire with our system, we need to train our model with similar data. We do this by manually annotating a set of images, which will be further described in the following.

## Instance Labeling

A subset of images are hand-annotated and used for both network refinement and to test the performance of the detection algorithm. The subset is labelled for instance segmentation so that pixels belonging to each object in the image is labelled separately with a polygon shape. Manually labelling images for instance segmentation is a time consuming and to ease the process we use a free web-based annotation tool *LabelMe* [29] to create polygons. Figure 5.2 shows the *LabelMe* interface where the user can separately draw polygons for each object as desired. When the user enters the interface a selected image from the repository will be displayed in the centre. The user may label object of interest by clicking on 'create polygon' and selecting control points along the boundary of the object. Each object is assigned a class by naming them appropriately and an image can contain several classes and objects.

The process of labeling is rather time-consuming when segmentation is desired, because the exact shape of each object has to be marked. Annotation using only a bounding box around objects is much faster. The simpler bounding box approach for labeling can be used for detection and classification, albeit with a different network.

## Data

The training data consist of images captured with our onboard RGB camera setup and additional images were acquired with a DSLR camera on separate trips. Images from internet sources are also added to the training data. All images are manually labelled using the previously mentioned technique.

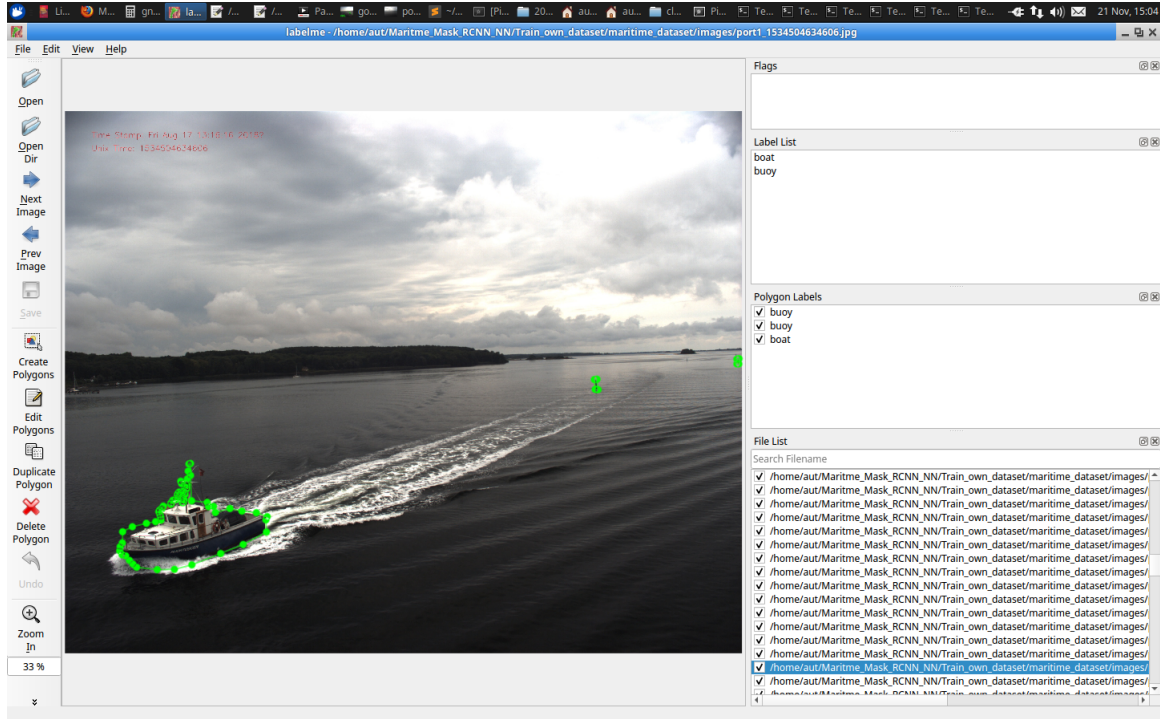


Figure 5.2: Screenshot of the *LabelMe* interface. The green polygons represent the boundary of the labelled pixels for a boat and two buoys respectively.

The validation set only consists of images from the onboard RGB camera setup, as we wish to evaluate the performance of the object detection on our on-board camera system. In summary, the labelled images for the dataset consists of:

Data source	Number of images
On-board RGB camera setup	330
On-board DSLR	179
Internet source	8
In total	517

The 517 images are annotated with two classes: buoy and ship. A total of 600 buoys and 639 ship instances are annotated across the dataset. Examples of the images labelled for instance segmentation are shown in Figure 5.3.

## 5.5 Training

We split the onboard RGB images so that 406 images are used for training and 111 images are used for validation. To produce additional training data we use the following data augmentation on each of the on-board RGB training images:

**Image rotation** Randomly rotating the images randomly between  $-25$  deg and  $25$  deg.

**Image flip** Flipping images horizontally (mirroring).

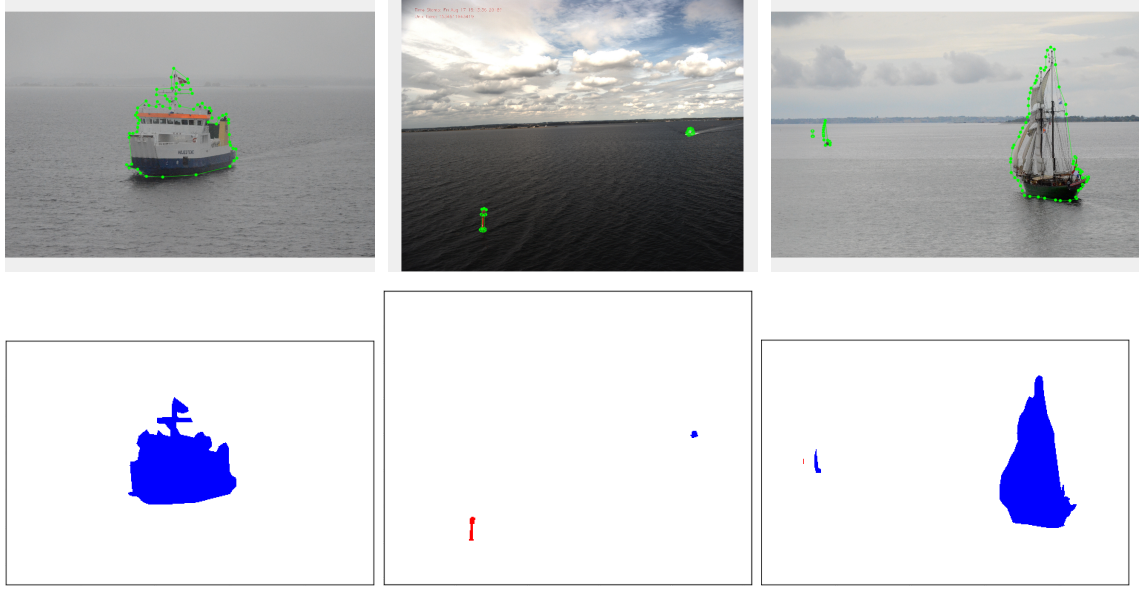


Figure 5.3: Sample images from our labelled dataset. The first row shows the images captured by our RGB cameras. The second row shows the masks of the labelled pixels, where the two classes are shown in red and blue respectively. Note that the instances within a class are labelled individually, despite they are visualized with the same colour.



Figure 5.4: Examples of data augmentation on the training set. The original training image is shown on the left and the next 4 images show the output after data augmentation.

**Image flipping and rotating** Combining the flipping and rotation.

**Noise addition** A colour replaces the image pixel for every 50 pixels.

The augmentation is visualized in Figure 5.4 and increases the dataset with an additional  $5 \times 406$  images. Each of these images are cropped into 16 regions in a  $4 \times 4$  grid. After this operation, the total increase of the dataset is  $16 \times 5 \times 406$  images, resulting in  $16 \times 5 \times 406 + 406 \times 5 = 34510$  images.

The Mask R-CNN has been pre-trained using the weights from the COCO dataset [20] and we fine-tune the network to detect the two classes provided in our training data: buoy and ship. The network was trained for 40 epochs on the first 4 layers (classificatory), then another 60 epochs for the rest of the layers and finally 80 epochs for the whole network. The learning rate was set to 0.0003 and

the momentum was 0.9. The total training time took around 24 hours on a GeForce GTX 1080 GPU.

## 5.6 Performance

We evaluate the performance of our network using the validation images consisting of only acquisitions from the on-board RGB camera system. With the above-mentioned training procedure, we obtain a mean average precision (mAP) of 62.74%. The 0.5-mAP is used which means that intersections of regions less than 50% are not included in the calculation.

For our study, we work with two stages of object detection. First of all, to detect and classify a relevant object in the image. Secondly, how accurately it is segmented. To discuss the results, we use the following terminology:

**True positive** Object present in the frame and detected.

**False positive** A detection occurs in the frame but without the presence of an object.

**True negative** Object not present in the frame and no detection occurs.

**False negative** Object present in the frame and not detected.

For our application, we need a good overall localization of the object in the image, but not necessarily a precise segmentation border around the object. Figure 5.5 shows sample predictions of the network and with visual inspection, we conclude that segmentation of the objects are acceptable in most cases where a true positive detection occurs.

We also wish to investigate to what extent the network is detecting the objects it is supposed to find, the occurrence of false positives and false classifications. To do this we note down the comparison of the reference (ground truth) annotations with the predictions provided by the network. The precision of the segmentation mask is omitted here, so it is only the object classification which is reflected in this part of the results. Note that our validation set consists of annotated images with one or more objects, but also images without objects are included in the set. Table 5.1 shows the results of the object detections and classifications. We consider the two object classes buoy and ship and divide the detections in a near and far field. The border between near and far is determined by human inspection and is therefore not a measurable distance.

The results show that the near-field object detections are working the best, 100% of the buoys and 100% ships are found and 0 buoys and ships misclassified. In the far field, only approximately 33% of the buoys and 66% of the ships are detected. Mis-classification is present at a far distance, where 1 buoy is detected as a ship, while 0 ships are detected as a buoy. A total of 6 buoys and 34 ships were detected in the images without actually being present. Note that these numbers were originally a bit higher, as our own ship is visible in some of our validation images, see examples in Figure 5.6. False detections caused by our own ship can be omitted as they can

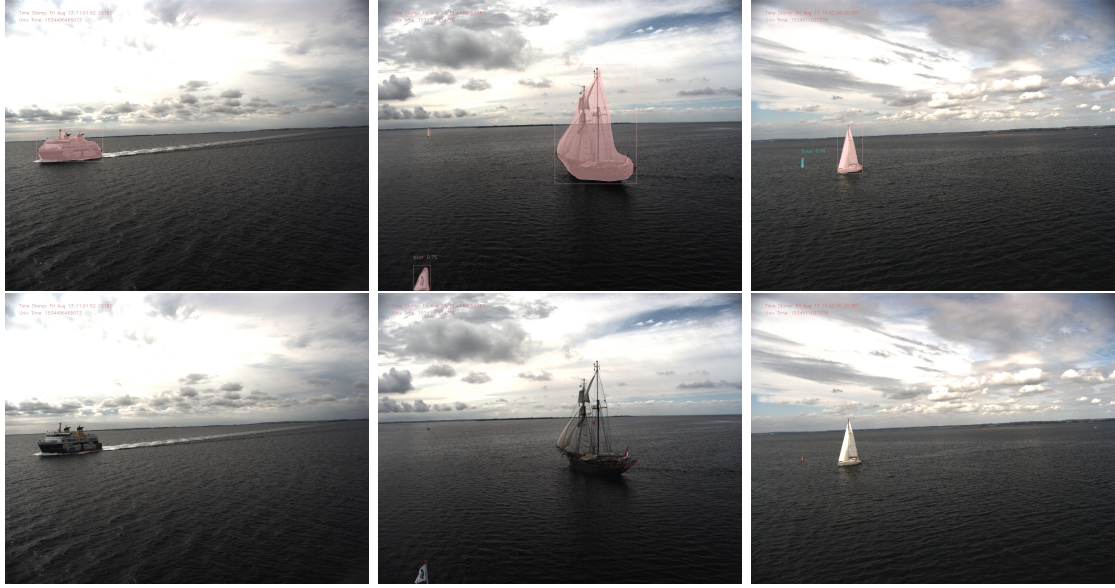


Figure 5.5: Example outputs from the Mask R-CNN network fine-tuned with our training data. The predicted instance segmentations are visualized on top of the input image.

		Detected				
		B	S	$\sim$ B	$\sim$ S	
Reference	near	B	47	0	0	-
		S	0	83	-	0
	far	B	27	1	54	-
		S	0	51	-	0
	none	$\sim$ B	6	-		-
		$\sim$ S		34	-	-

Table 5.1: Performance of the object detection. The detected objects are compared to the reference objects from the validation set. The number of detections (and no detections indicated with  $\sim$ ) is noted for the two types of objects: buoy (B) and ship (S). The buoy and ship detections are divided into two categories: near and far.





Figure 5.6: Examples of false positives. Some origin from sea reflections and others from clouds. Combination with classical methods could be one way to avoid such artifacts.

easily be identified as false positives since they will occur at a fixed location in all images.

The remaining part of the false positives are most often detections on the water where a piece of land far away is detected as a ship or in the region above the horizon line, where clouds are detected as ships. Figure 5.6 shows examples of the false positive detections. False positives in the cloud region could be removed by detecting the horizon in the image. Additionally, it would be an advantage to add more object classes to fully cover the objects present on the water in our image set.

The false negatives shown in Table 5.1 appear in particular when objects are far away. In particular, these missed detection of objects at long distance is related to the pixel area the object has in the images. Figures 5.7 and 5.8 show histograms of missed detections, shown in blue color, and of detections, shown in red color, as function of pixel area occupied by the object. All objects larger than 2500 pixels are detected but are not shown in these histograms. This classification behaviour is a consequence of the choice of network and of its training. The Figures also show that missed detection happens in a few cases for objects of larger area in the image. Object tracking and more robust classification methods could improve the detection, but it should be noted that approaching objects are eventually detected as distance to own ship decreases.

Other methods that are specialized to detect objects of very small size in images [9], could detect a rib boat already when it occupied 5 pixels in an image and could estimate the category when occupying  $38 \times 4$  pixels.

Using object tracking and fusion with information from NIR and WLIR imaging could provide additional suppression of artifacts, but this could not be included in the scope of this study.

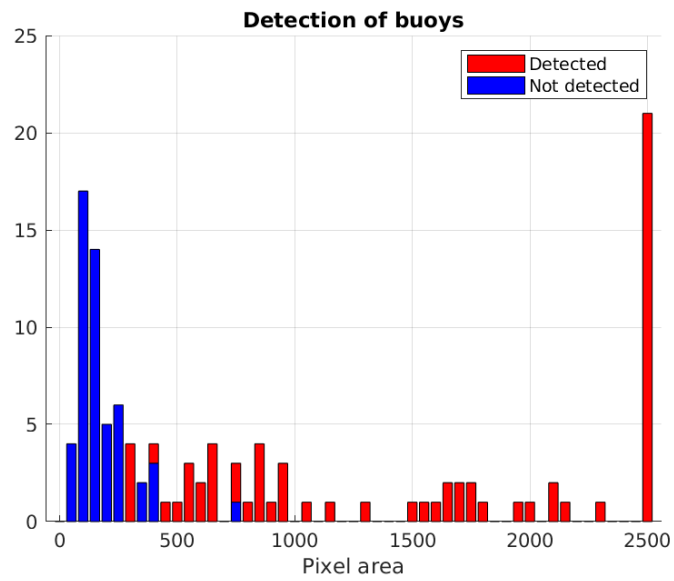


Figure 5.7: Histogram of pixel area versus buoy detections.

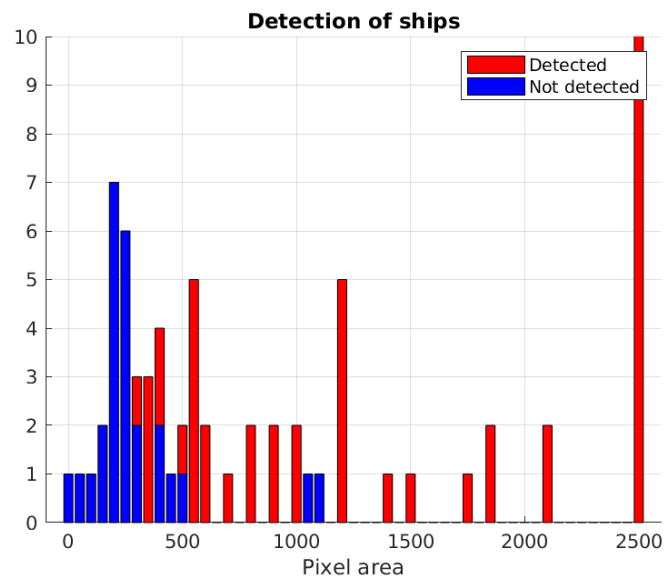


Figure 5.8: Histogram of pixel area versus ship detections.

# Chapter 6

## Results

This chapter compares the human outlook, made by assessing the fixations determined by the eye-tracking system, with object classifications made using RGB images.

Comparison between the electronic systems outlook capabilities and the human counterpart are hence done by looking at the instant of first observations of a given object. The eye-tracking software gives an indication of fixation on an object when the human lookout has been gazing at it for a certain length of time. This time is compared to the timestamp that the Mask R-CNN indicates its first detection and classification of the object. Figure 6.1 shows a snapshot of of eye-tracking. The right part shows what the lookout is focusing on. The yellow line on this shows that the eyes wanders around, which is normal. Fixation is indicated by the red circle. The Electronic Outlook is illustrated in Figure 6.2.



Figure 6.1: Eye-tracking of the manual look-outs fixations. Left: Forward facing camera used as reference in the analysis. Right: Eye-tracking result. The yellow spot surrounded by a thin red line indicates fixation on an object.

### 6.1 Temporal Comparison

This section presents an analysis of the time-wise differences between the electronic lookout system and the human counterpart. This is achieved by timestamping the detections of objects observed by the electronic lookout and comparing them with fixations captured by the eye-tracking system. A comparison is done by examining



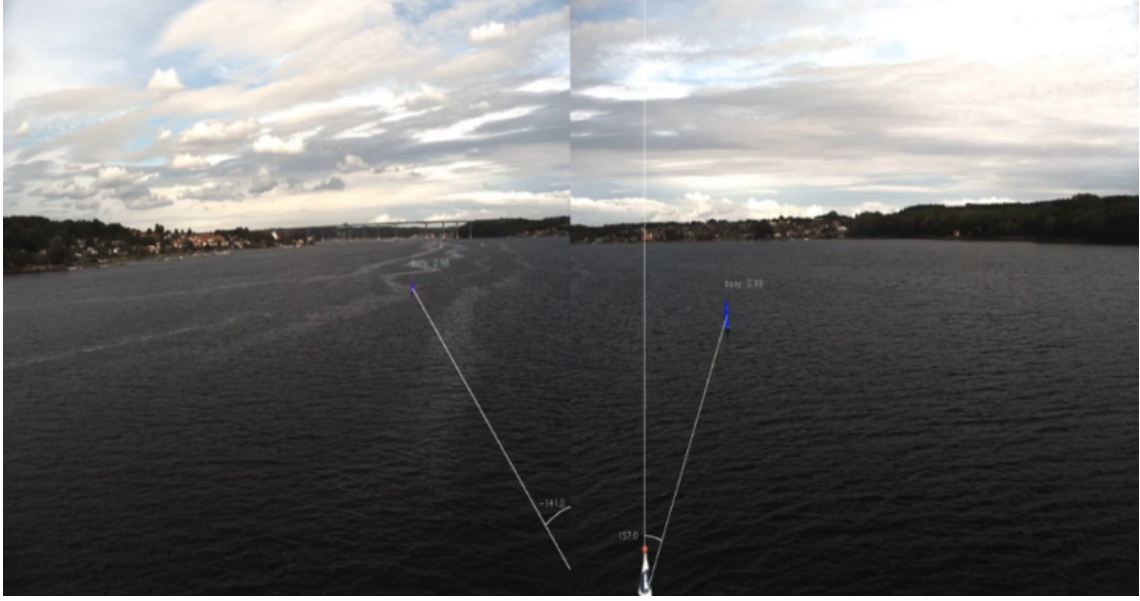


Figure 6.2: Object detection and classification on two RGB images are shown by highlighting the detected object and showing the bearing to detected objects.

the difference

$$\Delta t_{obs} = t_{HO} - t_{EO} \quad (6.1)$$

where  $t_{HO}$  is the time that the eye-tracking system indicates the first fixation on an object, and  $t_{EO}$  is the time that the electronic outlook first detects and classifies the same object. Figure 6.3 shows a histogram of  $\Delta t_{obs}$ . Figure 6.4 shows the

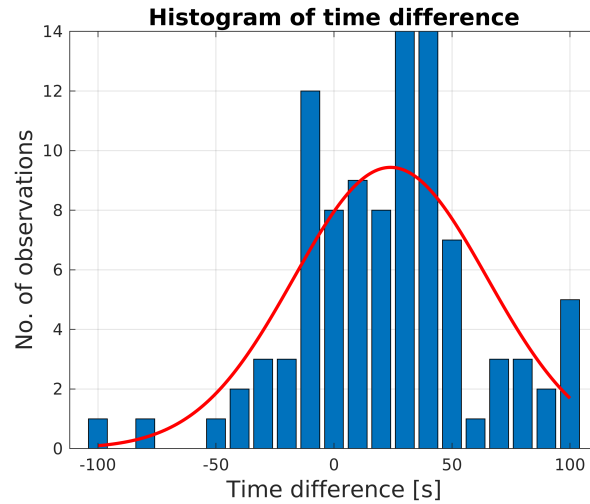


Figure 6.3: Histogram of time differences between observations done by the human lookout and the electronic lookout (calculated by (6.1)). The imposed normal distribution has the following parameters:  $\mu = 23.9$  s and  $\sigma = 41.0$  s. This means that electronic outlook classifies objects earlier than the human eye fixation by 24 seconds in average.

time difference  $\Delta t_{obs}$  histogram for ships and buoys separately. A positive value of

time difference means that electronic outlook classifies an object earlier than the navigator has a fixation on it.

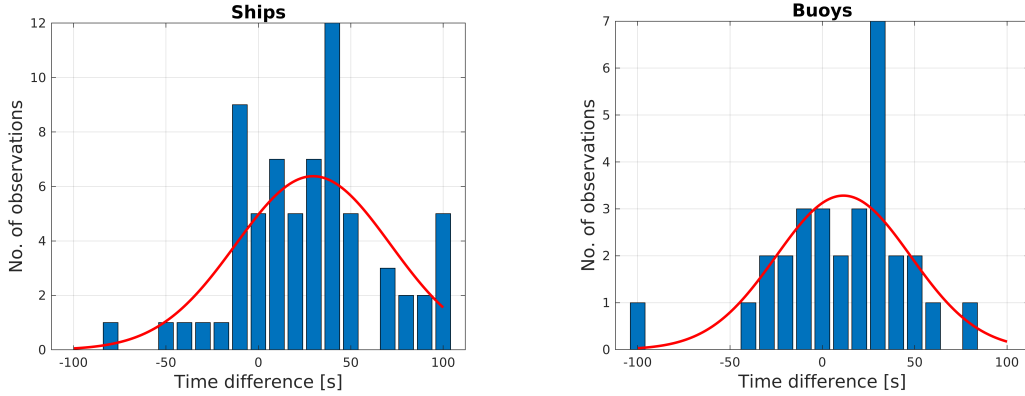


Figure 6.4: Histogram of time differences between observations done by the human lookout and the electronic lookout (calculated by (6.1)). In mean, the electronic outlook detects and classifies objects 30 s faster for ships and 11 s for buoys, compared to the human eye fixation. It is expected that the negative outliers could be avoided by improving the neural network.

The time elapsed between the instant of detection of an object and the instant when this object passes behind the RGB camera's field of view is defined as the time to react. Two time differences are defined to analyse this characteristic,

$$\Delta t_{HO} = t_{pass} - t_{HO} \quad (6.2)$$

$$\Delta t_{EO} = t_{pass} - t_{EO} \quad (6.3)$$

where  $t_{pass}$  is defined as the time instant when the object passes behind the RGB cameras' field of view.

Figure 6.5 shows a scatter plot of  $\Delta t_{HO}$  vs.  $\Delta t_{EO}$ . It is seen that electronic outlook allows more time to react. Figure 6.6 shows the same results but zoomed in at the range 0 – 200 s before passing own vessel. This range covers the majority of the measurements.

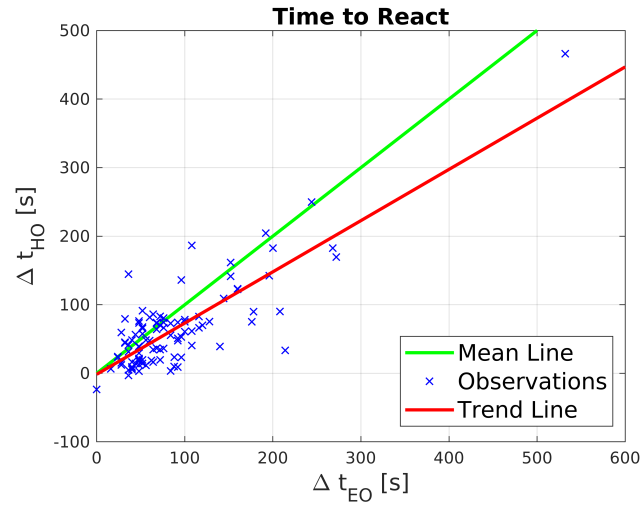


Figure 6.5: Scatter diagram of time to react. The trend line shows that time to react is longer with electronic outlook than time to react from a fixation.

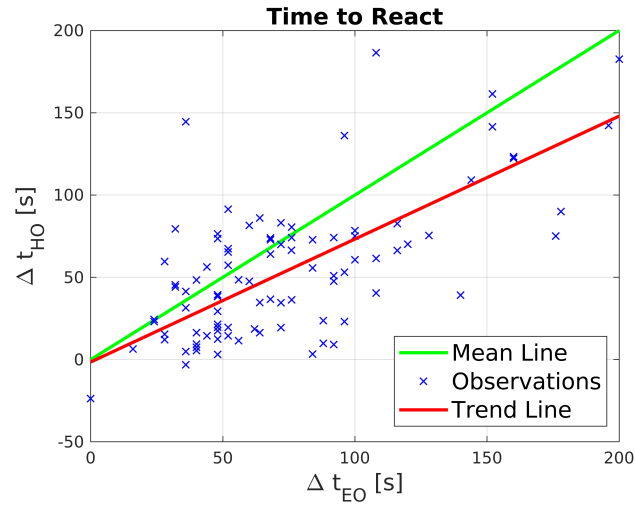


Figure 6.6: Scatter diagram of time to react. The plot shows the range 0 – 200 s. Again, the trend line shows that time to react is longer with electronic outlook than time to react from a fixation.

# Chapter 7

## Discussion

Since the ship has Radar and AIS sensors on board, the detection of objects that are visible to Radar or have AIS transmitters installed, could be done quite accurately. However, several objects are not visible on Radar, such as leisure surf borders and sea kayaks, boats without Radar reflector and AIS transmitter, and even containers that were accidentally dropped over board. Electronic outlook with object classification is therefore very important for the ship so that it acts in a safe manner also when non Radar detectable objects are in the area. Thus, a combination of object positions from these sensors and the Mask R-CNN architecture could increase the performance and the results. An example of this is by using the detected objects positions from the radar as possible region proposals in the network.

Further results will therefore fuse on-board radar and AIS information to improve the performance of the vision system. This will require calibration that enables Radar and AIS data to be studied from their respective coordinate systems into e.g. the pixel-coordinates of the input images to the CNN. This data could be used for region proposal in the network and be particularly useful in situations with reduced visibility of the cameras.

### Coverage of this analysis

Some kinds of behaviour are related to look-out, which are not captured by only observing the areas of fixtures with eye tracking glasses, but require further interpretation:

- General visual observation (watching) of nothing in particular, but often focused on the direction of the vessel and abeam/passed in relation to the progression of the navigation.
- Exogenous-oriented attention in relation to above item 1 – something turns up. This can include comparison or verification with information from instruments e.g. radar or AIS.
- Endogenous-driven observation of objects from other sources – for instance sea charts (buoys), radar and AIS (vessels) – expected to be observable.

Such interpretation of the situation, which is part of a situational awareness scenario, was not within the scope of this study.

Repeated observations to determine if there is a risk of collision and in this connection take countermeasures is visible in the eye-tracking measurements, and so is repeated observations to determine if countermeasures have the desired effect, but such awareness behaviours of the navigator are also outside the scope of the object detection and classification presented in this report.

### **Electronic outlook as decision support**

Look-out is just one among several tasks of the navigator on the bridge. Other tasks include: Observation of the condition of engines and systems; Handling of cargo and passengers; Safety-related routines; Communication internally on board the vessel and with external parties; Management of staff and other administrative tasks; QA and documentation tasks; Handling of safety-critical situations on board.

With several other tasks to care for, which might sometimes distract the navigator, it is believed that electronic outlook could serve as a *fifth sense* decision support for the navigator and perhaps make it possible to have temporally unmanned bridge in conditions with little to no other traffic.

# Chapter 8

## Conclusions

This study compared human outlook with electronic. Using instance of fixation of eye-tracking glasses with instance of electronic outlook by cameras and mask-RCNN classification, the study provided statistics for a comparison of instant of detection/fixation and time to object passes out of camera's field of view, which is close to passage of own ship and hence one of the essential parameters for object detection.

The performance of the Mask-RCNN was evaluated on the validation set of annotated RGB images. Object detection performance showed a satisfactory detection probability for objects larger than 400-500 pixels in an image, a quantification that is useful for camera system design for electronic outlook.

Some outliers were found to exist in form of false detections. A single instance of missed detections was also found in the validation data. Robustification of the classifiers will be needed to obtain the required dependability of electronic outlook and is a topic of further research, and techniques could include object tracking and combination of classical image processing with the deep learning methods for classification.

The situational awareness elements in a comparison were not covered but will be the subject of further research.

The main findings were:

- Eye-tracking glasses were found useful to show fixations on objects at sea in daylight conditions.
- The computer-vision algorithms detects objects in parallel, the human does so sequentially, and the computer classifies objects in average 24 sec faster than the navigator has a fixation on the object. The deep learning algorithm trained in this study should, however, be improved to achieve better performance in some situations.
- The time between object detection and passage of own ship is adequate for making navigation decisions with both human and electronic outlook.
- Low-light conditions (dusk and night) are effectively dealt with by Long Wave InfraRed (LWIR) camera technology. LWIR shows objects as equally visible at day and night.

- Colour information from cameras is necessary to assist decision support and electronic navigation.
- A system for electronic outlook should employ sensor and data fusion with radar, AIS and ECDIS.
- Decision support based on electronic outlook should include object tracking and situation awareness techniques.
- Quality assurance and approval of machine learning algorithms for object classification at sea has unsolved issues. A standard vocabulary ought be available for objects at sea, and publicly available databases with annotated images from traffic in both open seas, near coast areas and rivers should be available in order for authorities to assess quality or approve navigation support based on machine learning methods.

## Acknowledgements

The project will like to acknowledge the collaboration from Owner, Master and Navigators on M/F Højestene, M/F Marstal and M/S Pernille for allowing us on board for eye-tracking measurements and to these ferries and to M/F Isefjord for collaborating by inviting the team on board for camera mapping of objects at sea where the respective ferries operate. The authors would also like to acknowledge the dedicated efforts made by present and former laboratory engineers and graduate students. Jakob Bang from the Danish Maritime Authority and Jens Peter Weiss Hartmann from the Danish Geo Data Authority have guided and advised during the study, and the study was made possible by funding from the Danish Maritime Foundation via DTU's Maritime Centre. The authors gratefully appreciate the enthusiastic collaboration with the above mentioned collaborators and advisers and acknowledge, in particular, the funding made available.

# Bibliography

- [1] (2018). *Tobii Pro Glasses 2, User's Manual*, 1.0.5 edition.
- [2] Aleem, S. (2017). *Analysis of shiphandlers' eye-gaze and simulation data for improvements in COVE-ITS system*. PhD thesis, Naval Postgraduate School,, Monterey, California.
- [3] Bjørneseth, F., Loraine, C., Dunlop, M., and Komandur, S. (2014). Towards an understanding of operator focus using eye-tracking in safety-critical maritime settings. In *Proc. Int. Conference on Human Factors in Ship Design & Operation*, Glasgow.
- [4] Blanke, M., Hansen, S., Maurin, A. L., Stets, J. D., Koester, T., Brøsted, J. E., Nykvist, N., and Bang, J. (2018). Outlook for Navigation – Comparing Human Performance with Robotic Solution. In *Proc. ICMASS 2018, 1st International Conference on Maritime Autonomous Surface Ship*, Busan, Republic of Korea.
- [5] Bloisi, D. D., Pennisi, A., and Iocchi, L. (2014). Background modeling in the maritime domain. *Machine vision and applications*, 25(5):1257–1269.
- [6] Bousetouane, F. and Morris, B. (2016). Fast cnn surveillance pipeline for fine-grained vessel classification and detection in maritime scenarios. In *Advanced Video and Signal Based Surveillance (AVSS), 2016 13th IEEE International Conference on*, pages 242–248. IEEE.
- [7] Can, T., Karali, A. O., and Aytac, T. (2011). Detection and tracking of sea-surface targets in infrared and visual band videos using the bag-of-features technique with scale-invariant feature transform. *Applied Optics*, 50(33):6302 – 6312.
- [8] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. Ieee.
- [9] Dulski, R., Milewski, S., Kastek, M., Trzaskawa, P., Szustakowski, M., Ciurapinski, W., and yczkowski, M. (2011). Detection of small surface vessels in near, medium and far infrared spectral bands. In *Proceedings of SPIE*, volume 8185, pages U1–U11, Prague.
- [10] Feflatyev, S., Goldgof, D., Shreve, M., and Lembke, C. (2012). Detection and tracking of ships in open sea with rapidly moving buoy-mounted camera system. *Ocean Engineering*, 54:1–12.



- [11] Forsman, F., Sjors, A., Dahlman, J., Falkmer, T., and Lee, H. C. (2012). Eye tracking during high speed navigation at sea. *Journal of Transportation Technologies*, 277.
- [12] Girshick, R. (2015). Fast r-cnn. *2015 IEEE International Conference on Computer Vision (ICCV)*.
- [13] Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*.
- [14] Hareide, O. and Ostnes, R. (2017a). Scan pattern for the maritime navigator. *Transnav – the International Journal on Marine Navigation and Safety of Sea Transportation*.
- [15] Hareide, O. S. and Ostnes, R. (2017b). Maritime usability study by analysing eye tracking data. *The Journal of Navigation*, 70(5):927–943.
- [16] He, K., Gkioxari, G., Dollar, P., and Girshick, R. (2017). Mask R-CNN. *2017 IEEE International Conference on Computer Vision (ICCV)*.
- [17] Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and Van de Weijer, J. (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press.
- [18] Leclerc, M., Tharmarasa, R., Florea, M. C., Boury-Brisset, A.-C., Kirubaranjan, T., and Duclos-Hindié, N. (2018). Ship classification using deep learning techniques for maritime target tracking. In *2018 21st International Conference on Information Fusion (FUSION)*, pages 737–744. IEEE.
- [19] Leira, F. S., Johansen, T. A., and Fossen, T. I. (2015). Automatic detection, classification and tracking of objects in the ocean surface from uavs using a thermal camera. In *Aerospace Conference, 2015 IEEE*, pages 1–10. IEEE.
- [20] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer.
- [21] Moreira, R. D. S., Ebecken, N. F. F., Alves, A. S., Livernet, F., and Campillo-Navetti, A. (2014). A survey on video detection and tracking of maritime vessels. *International Journal of Recent Research and Applied Studies*, 20(1).
- [22] Muczynski, B., Gucma, M., Bilewski, M., and Zalewski, P. (2013). Using eye tracking data for evaluation and improvement of training process on ship’s navigational bridge simulator. *Maritime University of Szczecin*, 33(105):75–78.
- [23] Pico, M., D., H., Bik, R., van der Wiel, S., and van Basten Batenburg, R. (2015). Enhancing situational awareness. a research about improvement of situational awareness on board of vessels. Technical report, . Rotterdam Mainport University of Applied Sciences.

- [24] Prasad, D. K., Rajan, D., Rachmawati, L., Rajabally, E., and Quek, C. (2017). Video processing from electro-optical sensors for object detection and tracking in a maritime environment: a survey. *IEEE Transactions on Intelligent Transportation Systems*, 18(8):1993–2016.
- [25] Ramírez, J. M. (2018). Detection and classification of objects at sea using computer vision. Master’s thesis, Technical University of Denmark, Department of Electrical Engineering, Automation and Control Group.
- [26] Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149.
- [27] Renganayagalu, S. and Komandur, S. (2013). Video support tools for training in maritime simulators. In *Proc. of the International Conference on Contemporary Ergonomics and Human Factors*, Cambridge UK.
- [28] Rodin, C. D. and Johansen, T. A. (2018). Detectability of objects at the sea surface in visible light and thermal camera images. In *Proceedings of OCEANS’2018*.
- [29] Russell, B. C., Torralba, A., Murphy, K. P., and Freeman, W. T. (2008). Labelme: A database and web-based tool for image annotation. *Int. J. Comput. Vision*, 77(1-3):157–173.
- [30] Salvador, E. V. (2018). Image analysis using machine learning methods of electronic outlook on board ships. Master’s thesis, Technical University of Denmark, Department of Electrical Engineering, Automation and Control Group.
- [31] Stadler, M. (1984). *Psychology of Sailing. The sea’s effects on mind and body*. Adlard Coles Ltd.
- [32] Szpak, Z. L. and Tapamo, J. R. (2011). Maritime surveillance: Tracking ships inside a dynamic background using a fast level-set. *Expert systems with applications*, 38(6):6669–6680.
- [33] van Westrenen, F. C. (1999). *The Maritime Pilot at Work. Evaluation and use of a time-to-boundary model of mental workload in human-machine systems*. Dr.ing. thesis, The Netherlands TRAIL Research School. TRAIL Thesis Series nr T99/2.
- [34] Zhang, Y., Li, Q.-Z., and Zang, F.-N. (2017). Ship detection for visual maritime surveillance from non-stationary platforms. *Ocean Engineering*, 141:53–63.